Interactive comment on “A pan-African Flood Forecasting System”
By V. Thiemig et al.

Anonymous Referee #2
Received and published: 13 July 2014

We thank the reviewer very much for dedicating his/her time to make this very sophisticated review, pointing out very many issues, giving us the possibility to clarify and improve our manuscript.

GENERAL COMMENTS This m/s evaluates the performance of a flood forecasting system over Africa. In itself, describing a new forecasting system and evaluating its performance without comparison with existing approaches or systems in my view does not merit publication in HESS. However, I do think there are good opportunities in this m/s to answer some important questions, e.g. to what extent (ECMWF) forecasts can increase forecasting skill, and whether a forecasting systems like this can add value to alternative early warning approaches. Currently it does not address such important questions yet, but with some additional analysis it could.

HESS holds a long history of publications related to the development and testing of novel flood forecasting approached and systems. This is commonly known and appreciated by researchers of this particular field. Here only a small selection of similar studies already published in HESS:

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: GloFAS – global ensemble streamflow forecasting and flood early warning, Hydrol. Earth Syst. Sci., 17, 1161-1175, doi:10.5194/hess-17-1161-2013, 2013.

Thielen, J., Bartholmes, J., Ramos, M.-H., and de Roo, A.: The European Flood Alert System – Part 1: Concept and development, Hydrol. Earth Syst. Sci., 13, 125-140, doi:10.5194/hess-13-125-2009, 2009.

Bartholmes, J. C., Thielen, J., Ramos, M. H., and Gentilini, S.: The european flood alert system EFAS – Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts, Hydrol. Earth Syst. Sci., 13, 141-153, doi:10.5194/hess-13-141-2009, 2009.

Leedal, D., Weerts, A. H., Smith, P. J., and Beven, K. J.: Application of data-based mechanistic modelling for flood forecasting at multiple locations in the Eden catchment in the National Flood Forecasting System (England and Wales), Hydrol. Earth Syst. Sci., 17, 177-185, doi:10.5194/hess-17-177-2013, 2013.

Jaun, S. and Ahrens, B.: Evaluation of a probabilistic hydrometeorological forecast system, Hydrol. Earth Syst. Sci., 13, 1031-1043, doi:10.5194/hess-13-1031-2009, 2009.

Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., Verseghy, D., Soulis, E. D., Caldwell, R., Evora, N., and Pellerin, P.: Development of the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale, Hydrol. Earth Syst. Sci., 11, 1279-1294, doi:10.5194/hess-11-1279-2007, 2007.

Further, the manuscript provides not only a description of the newly developed African Flood Forecasting System (AFFS), but also presents and evaluates the systems performance. This is done not for a couple of selected flood events and specific catchments, but for a whole year and for a whole continent – which on its own is an enormous effort that has not been done so far by any study known to us.

The fact that the system has not been compared to others in this manuscript is excused by the scope and complexity of this present manuscript. Adding a major part to it, would not necessarily add to the value of the paper. Our intention was to introduce a new system and to give a first evaluation on its performance. More on the issue of comparison to other system later during this document, but please note that none of the already published studies cited above hold such a comparison.

Concluding on this point, we consider HESS as the right source for publishing this particular manuscript on AFFS as it contributes to the exploration of possible flood forecasting solutions and complements the already existing collection as presented above.

I have the following main criticisms with the m/s as it stands:

1) The analysis presented does not consider the source of forecast skill. Previous studies have established that much of the skill in streamflow forecasts comes from initial catchment conditions rather than weather forecasts, so an evaluation like this should tease those two apart. A simple way of doing that would be to compared the forecasts derived with ECMWF forecasts to those derived with a 'traditional' ensemble approach where forcings are drawn from the historic record. That would also be a more appropriate benchmark than the seasonal mean – indeed a simple persistence assumption would probably already be a more appropriate assumption than the seasonal mean.

Yes, this is indeed a very good suggestion of the reviewer. However, we feel it is beyond the scope of this paper to establish the sources of predictability as we focus here on establishing that there is skill as well as on introducing the concept of AFFS. A detailed study on establishing the sources of the forecast skill will be done as a next step together with further model improvements. We included, however, some information on the "conventional ESP" (See detailed reply to 5575-5 below).

2) Another (related) problem with interpreting the results for longer rivers and justifying the conclusion that ''lead-time could easily be extended up to 15 days by using the ECMWF-ENS" (5579-4) is that there is no analysis of what part of skill and

lead time derives from the delay from river routing (i.e. the early warning provided by the time it takes for a flood pulse to travel downstream) versus the real contribution from ECMWF forecasts.

The choice of wording we used was misleading, we meant to say that the _flood forecasts_ *could be extended up to 15 days. We changed that in the manuscript. How much the extension of 5 days actually contribute is – as mentioned by the reviewer – a matter of catchment size and other influencing factors. However, as already mentioned above to establish the source of predictability was unfortunately beyond the scope of this study but will be subject of future research.*

3) Limiting the evaluation to the 'flood-intense year 2003' (p 5562 l 28) could potentially skew the verification metrics. Ideally, evaluation would be extended to a more representative time period but I appreciate that there may be logistical challenges with that. Nonetheless it is a potentially important caveat that needs some discussion.

Yes, we agree. The ideal validation for AFFS would span over the whole time period that is covered by both, meteorological EPS and ground observations. However, as the reviewer already indicated, there are technical constrains given by the size of raw and resulting data as well as time needed for computation of the forecasts and the evaluation of the results.

Many studies already published (see below) focus on the calculation and evaluation of flood forecasts for a single or a number of selected events and catchments:

Numerical simulations of the 12-13 August 2002 flooding event in eastern Germany (2004) Quarterly Journal of the Royal Meteorological Society, 130 (600 PART A), pp. 1921-1940. ***DOI:*** *10.1256/qj.03.152*

Anderson, M.R., Rowe, C.M., Radell, D.B., McCormick, J.R. Flash flooding during a severe drought: A case study of the 2002 Ogallala, NE event (2004) Bulletin of the American Meteorological Society, pp. 2639-2641. ***ISSN:*** *00030007*

Blöschl, G., Reszler, C., Komma, J. Hydrological flooding forecasts for the Kamp - Experiences with the occurrences for 2006 and 2007 [Hydrologische Hochwasservorhersage für den Kamp - Erfahrungenmit den Ereignissen 2006 und 2007] (2008) Osterreichische Wasser- und Abfallwirtschaft, 60 (3-4), pp. a13-a18. ***DOI:*** *10.1007/s00506-008-0157-y*

Smith, J.A., Baeck, M.L., Villarini, G., Wright, D.B., Krajewski, W. Extreme flood response: The june 2008 flooding in Iowa (2013) Journal of Hydrometeorology, 14 (6), pp. 1810-1825. ***DOI:*** *10.1175/JHM-D-12-0191.1*

Hopson, T.M., Webster, P.J. A 1-10-day ensemble forecasting scheme for the major river basins of Bangladesh: Forecasting severe floods of 2003-07 (2010) Journal of Hydrometeorology, 11 (3), pp. 618-641. ***DOI:*** *10.1175/2009JHM1006.1*

Ramos, M.-H., Thielen, J., De Roo, A. Ensemble Hydrological Forecasting and Alert with the European Flood Alert System (EFAS): Case of the Danube Basin Floods in August 2005 (2013) Practical Applications in Engineering, 4, pp. 47-61. ***DOI:*** *10.1002/9781118557792.ch5*

He, Y., Wetterhall, F., Bao, H., Cloke, H., Li, Z., Pappenberger, F., Hu, Y., Manful, D., Huang, Y. Ensemble forecasting using TIGGE for the July-September 2008 floods in the Upper Huai catchment: A case study (2010) Atmospheric Science Letters, 11 (2), pp. 132-138. ***DOI:*** *10.1002/asl.270*

Ushiyama, T., Sayama, T., Tatebe, Y., Fujiokacc, S., Fukami, K. Numerical simulation of 2010 Pakistan flood in the Kabul river basin by using lagged ensemble rainfall forecasting (2014) Journal of Hydrometeorology, 15 (1), pp. 193-211. ***DOI:*** *10.1175/JHM-D-13-011.1*

In the case of AFFS, however, to limit the verification to a single or a couple of selected flood events or some selected target catchments would not have been satisfying to us, as AFFS is a continental flood forecasting system and hence needed to be verified as such. Yet, to calculate a continuous probabilistic medium-ranged flood forecast for a whole continent, in a 0.1° resolution, for a longer time period entails non-trivial challenges. For example, the size of the ECMWF-ENS (50 ensembles, 10 day forecast, 8 meteorological variables) for one forecasting day is around 5.4 Gb and around 2 Tb for one year. On top, the same amount of disk space was needed to store the resulting hydrological predictions, including threshold exceedance maps. The computation time, however, is a function of IT sources available and can vary accordingly. In our case we were quite well-equipped with access to a 48-core cluster LINUX machine. Despite that, the time needed for running all the scripts (from scraping input data from ECMWF-FTP to the processing of the hydrological predictions into flood forecasts) added up to a couple of month – excluding the time required for the analysis of the results.

With these constraints in mind, we performed a pre-analysis based on information collected from different disaster data basis between 2000 and 2012 and decided to focus the verification for the year 2003 as it was the year with the most flood events reported.

4) There should be more discussion of the 'optimization' of the model (5564-11) and the extent to which this may have compromised independent forecast verification. For example, did you produce 1 universal parameter set or 36 different ones (5568-14)? How did you obtain parameters for other catchments and what are the implications for the forecasts? If I understand Section 5.1 correctly, the calibration and validation were not at all independent.

We thank the reviewer for this comment and that we can clarify the 'optimization' section which might be misleading. In sentence 5564-11 we might have given the impression that the calibration period is linked to (when, where and with which magnitude) flood events and therefore not independent. However, the calibration period is chosen only dependent on the observed discharge data availability and is totally not related to flood events. In Section 5.1 *where we describe figure 6 the calibration period (2004-2008) is totally independent of the validation period (1998-2003). We rephrased the sentence in 5564-11 with:*

5564-11: "Discharge observations are required for the optimization of LISFLOOD and information about floods, in particular on *when*, *where* and with which *magnitude* a flood event has happened, is required for the verification of the performance of AFFS. Therefore, discharge observations and information retrieved from various flood archives were employed as hydrological reference data."

For the calibration procedure we obtained 36 unique 'best' parameter sets. For catchments lacking sufficient data for model calibration the default values without calibration were used for the model parameters. We will add an extra column with the default values in Table 2. The purpose of calibration is to obtain a model parameter set which capture the behavior of the catchment and therefore should result in better performance of the model. Using the default parameter values should therefore result in a weaker model performance and forecast when the conditions (area, model input accuracy) of the catchments are equal. We adapted the text of page 5568 to make the optimization description more clear.

5568-14: "Meteorological variables were obtained from the ERA-Interim and ECMWF-EPS fields (Simmons et al., 2007) from the European Centre for Medium-range Weather Forecasts (ECMWF). Parameters related to groundwater response, infiltration, groundwater losses, channel routing and reservoir operating rules were determined through model calibration.
The pan-African set-up was calibrated for each individual of the 36 sub-catchments (see green dots in Fig. 1e), corresponding to 11 hydrological basins, over a time period of five years (2004–2008; 2003 used as warm-up). To drive LISFLOOD in the calibration procedure, the ERA-Interim precipitation which was corrected using the Global Precipitation Climatology Project (GPCP) dataset from the ECMWF was used. This is done because Balsamo et al. (2010)..."

The benefits of using KGE' over KGE or Nash–Sutcliff Efficiency are discussed by Gupta et al. (2009) and demonstrated by Thiemig et al. (2013).
After the calibration, a unique "best" parameter set was obtained. For catchments lacking sufficient data for model calibration default values without calibration were used for the model parameters (Table 2).

SPECIFIC COMMENTS

Abstract: I could not initially work out what "Save flooding" referred to. Suggest 'case study for flooding in the Save River' (if you keep this part of the m/s)

We agree that this is misleading i.e. not clear what is meant. Therefore we changed it to:

The case study for the flood event in March 2003 in the Sabi Basin (Zimbabwe) illustrated…

The name of the basin that we referred to as Save was changed to Sabi as one of the reviewers kindly pointed out that this basin is the more commonly known in Africa under this name.

L 21) What is the H for?

HEPS = Hydrological Ensemble Prediction System
All acronyms are now defined at first mentioning. We also included a glossary at the end of the manuscript.

L 26) Table 1 includes at least one system that is not a HEPS (the Bureau of Meteorology system, which uses a Bayesian approach). Overall this information would be more valuable if characterised or categorised and discussed in the text. For example, which systems use weather forecasts and which take a (conditional) sample from the historic record; how is initial state considered in the different systems; which consider river routing. This sort of context is necessary to explain the innovation in this m/s.

Table 1 was updated (see below) and made sure it only contains HEPS. A detailed review *of all available systems* is completely beyond the scope of this paper. However, there are two forthcoming publications which will contain considerable detail and the interested reader is referred to those:

1. Flood Forecasting: a Global Perspective, Thomas Adams and Tom Pagano (Editors), Anticipated Publication date: May 2015
2. HEPEX book

Both are in their final stage and we hope that we can link to them to give the reader the possibility to access more information on this topic.

Forecast centre / System name	**Provider**	**Domain**	**Status**	**Reference**
AIGA-Ensemble	IRSTEA	Southern France, to be extended to mainland France	T	Lavabre and Gregoris (2006), Javelle et al. (2009, 2012)
CHROME	SCHAPI and Météo-France	Gardon d'Anduze, Ardeche and Ceze river (France)	PO	
European Flood Awareness System (EFAS)	European Commission - Copernicus program	Europe	O	www.efas.eu

FEWS Scotland	Scottish Flood Forecasting Service	Scotland	O	Cranston and Tavendale (2012), Werner at al. (2009, 2013)
Flood Early Warning System for the Po River and the Emilia Romagna Region (FEWSOO/ER)	ARPA Emilia Romagna – Italy	Po, Reno and Romagnoli river (Northern Italy)	O	http://www.deltares.nl/en/project/101490/flood-forecasting-system-river-po-italy/873016; Werner et al. (2013)
Global Flood Awareness System (GloFAS)	JRC/ECMWF	Global	PO	www.efas.eu, Alfieri et al. (2013)
Hydrologic Ensemble Forecasting Service (HEFS)	US National Weather Service	USA	O	Demargne et al. (2014); http://www.nws.noaa.gov/oh/XEFS/; http://www.cbrfc.noaa.gov/devel/hefs/; http://www.erh.noaa.gov/mmefs/index.php
Hugo	Bayerisches Landesamt für Umwelt	Bavaria (south-east Germany)	O	http://ksh.fgg.uni-lj.si/bled2008/cd_2008/01_Hydrological%20forecasting/182_Hangen-Brodersen.pdf
Hydrological warning system for Norway (HWN)	Norwegian Water Resources and Energy Directorate, Hydrology Department	Continental Norway	O	http://www.nve.no/en/Floods-and-landslides/Flood-forecasting-system/; http://www.varsom.no/Flom/
IFKIS-Sihl / IFKIS-Ticino	WSL	Sihl and Ticino river (Switzerland)	O	Romang et al. (2011)
LARSIM Moselle and Rhineland-Palatinate	Landesamt für Umwelt, Wasserwirtschaft und Gewerbeaufsicht Rhineland Palatinate (Germany)	Moselle river (France, Luxembourg, Germany) and federal state of Rhineland-Palatinate (Germany)	T	http://www.hydrology.uni-freiburg.de/publika/band22.html; http://www.watlab.be/en/events/files/International%20Symposium%20on%20Innovations%20in%20Flood%20Forecasting%20Systems%20/gerlinger
AquaLog MESP (monthly ensembled streamflow prediction)	Czech Hydrometeorological Institute	Czech Republic	O	Březková et al. (2007); http://www.nve.no/PageFiles/6652/poster08FIN.pdf

Meteorological Model-based Ensemble Forecast System (MMEFS)	NOAA/NWS	Most of the eastern US	O	http://www.erh.noaa.gov/mmefs/index.php;
Novel Flood Warning and Risk Assessment System (NEWS)	NEWS	Upper Huai (China)	E	http://news.nmpi.net/
Pilot EPS Rijnland (PER)	UNESCO-IHE	Rhine Delta (The Netherlands)	PO	http://meetingorganizer.copernicus.org/EGU2013/EGU2013-9451.pdf
PREDICTOR	EDF	France	O	http://hepex.irstea.fr/wp-content/uploads/2013/12/EDFmathevet-9-2011.pdf; Mathevet et al. (2009)
RWsOS Rivers	Rijkswaterstaat	Rhine and Meuse river	O	http://www.lthe.fr/PagePerso//chardon/doc/chardon_EGU_2012.pdf
AquaLog/Hydrog SESP (short-term ensemble prediction)	Czech Hydrometeorological Institute	Czech Republic	O	Březková et al. (2007); http://www.nve.no/PageFiles/6652/poster08FIN.pdf
Loire and Allier Forecasting System (SPC-LCI)	Service de Prévision des Crues Loire-Cher-Indre and Service de Prévision des Crues Allier	Loire and Allier river (France)	O	-
Seasonal Streamflow Forecast (SSF)	Bureau of Meteorology	Australia	O	Laugesen et al. (2011); http://www.bom.gov.au/water/ssf/
WAVOS, FEWS (combination of two forecast systems): WAVOS focuses on hydrodynamics, FEWS focuses on hydrology.	German Federal Institute of Hydrology (BfG)	Germany, German Federal Waterways	O	http://www.bafg.de/DE/08_Ref/M2/04_Vorhersagen/Einsatzgebiete/einsatzgebiete_node.html;jsessionid=59851CB288BC7DC67CC234B3BB0B4EDE.live1042#Start
Water Problems Institute of Russian Academy of Sciences (WPI RAS)	Water Problems Institute of Russian Academy of Sciences	Vyatka, Sosna and Seim river (European part of Russia)	O	Kuchment and Gelman (2007, 2009)
Watershed Simulation and	Finnish Environment	Finland and border crossing	O	http://www.syke.fi/download/noname/%7B4D2F88B9-21F6-

Forecasting System WSFS	Institute	rivers	4ED5-AEB7-C1AD30A94D70%7D/32817

5564-18) How did you separate flash floods from medium- to large-scale floods? (i.e., what criterion, what method, what data)

Our reference database of reported flood events excludes flash floods (page 5564, l. 18). The exclusion was mainly based on the explicit statement given by the disaster databases (Dartmouth Flood Observatory; Emergency Event Database EM-DAT). In a couple of cases, where the flood was not reported by Dartmouth Flood Observatory or Emergency Event Database EM-DAT, but by the NASA Earth Observatory or Reliefweb (which usually don't specify the specific type of riverine flooding), either the duration of flooding and the indicated inundated extend were used to filter flash floods out. Hence, if the flood duration was only one or two days, or the extend local it was presumed to be a flash flood and the event was excluded. The resulting flood events are medium- to large scale events (p. 5564, l19) by their extent and duration.

Section 3.2) Please provide a table with all model input and forcing data sources and URL download links, so that future researchers might attempt to (more or less) replicate your experiment.

The authors don't agree with the idea to provide a table with all model input as this will be an inexhaustible list. As stated in the manuscript (line 5567-25) a detailed description of the input maps can be found in Bodis (2009). Together with the manual (Burek et al., 2013) it should give the reader a clear overview of the static input data. For all the meteorological data future researchers should contact the ECMWF as all forcing data is retrieved from the ECMWF. We changed some lines to make this clearer:

5567-25: "A list of all the required input maps is given in Burek et al. (2013) and a more detailed description of the source of the input maps for Africa is specified by Bodis (2009)."

5568-10: "Meteorological variables were obtained from the ERA-Interim and ECMWF-EPS fields (Simmons et al., 2007) from the European Centre for Medium-range Weather Forecasts (ECMWF).

5568-17: "To drive LISFLOOD in the calibration procedure, the ERA-Interim precipitation which was corrected using the Global Precipitation Climatology Project (GPCP) dataset from the ECMWF was used."

5570-15) In the discussion, please address how changing characteristics of the GPCP and forecast data over time (e.g. analysis uncertainty and forecast quality resp.) may influence the performance of your system. In particular, how representative is the ECMWF forecast skill for 2003 for the skill in 2014?

As we don't have the chance to answer this question based on evaluating AFFS for 2014 (as we are lacking observations and the work load for running another year of forecast *would be tremendously large as already indicated under point 3 of the “main criticisms”), we have to fall back on past experience.*

Pappenberger et al. published a study in 2011 (DOI: 10.1002/hyp.7772) a study, in which the impact of weather forecasts improvements on large scale hydrology was analyzed. “Although results show that improved meteorological predictions clearly contribute to better hydrological forecasts, a dominant influence of natural variability and observed meteorological data can also be observed. Sudden increase in skill apparent in meteorological models does not consequently lead to similar improvements in the hydrological modelling, where other influences may play equally important role. Over the 10-year period considered in this study, the EFAS system forecast skill improved by 2.2 days on average and has a theoretical skilfulness for a lead time of 30 days.”

Based on this the discussion was extended accordingly:

… Additionally, a cross-comparison study of AFFS with other global forecasting systems is necessary to gain a deeper understanding on the particular strengths and limitations of AFFS, as well as to examine issues such as whether there is a necessity for a hydrological model, or the detail of output products required to be useful for the end-users. This will therefore be the focus of future research. Also, based on information from Pappenberger et al. (2011) the performance of AFFS might be even better in recent years, as a consequence of the continuous improvement of the quality of meteorological data used as input to AFFS. However, also this needs to be addressed in future research. *The HEPEX initiative (www.hepex.org) and the recently launched Global Flood Partnership http://portal.gdacs.org/Global-Flood-Partnership) will be explored as a possibility for further testing of AFFS in research and experimental real-time mode.*

5573-8) How were these areas delineated? They look like they were done by hand, which makes me wonder about interpretation bias.

The delineation of the affected areas was done based on the threshold exceedance maps, which show in each 0.1° pixel the number of ensemble members exceeding the 2- and 10 year return period threshold. To decrease the noise which might happen, we accumulated 5 and 10 days of threshold exceedance maps and draw (with ArcGIS) a line around the areas that showed a large number of ensembles exceeding the critical hydrological thresholds.

As the delineation in Figure 4 was only used to indicate the location of forecasted flood events (also noted like that in the caption of Figure 4) and not for some analysis comparing the forecasted with reported flood extend and location, we don’t face the problem of interpretation bias.

5573-16) That is a very ad hoc interpretation. If you knew that was true for those data and not for other data, you should not have used them in the first place?

We agree that “questionable” is a not well-suited word in this sentence. The relevant station is located downstream the large Niger delta swamp and therefore the observed discharge *has a different characteristics compared to the stations upstream the delta, which doesn't have to per definition questionable. We rephrased the sentence from:*

5573-14: "Poorer hydrological performances (KGE'<0.5) are clustered in smaller tributaries in the arid area of South Africa and in a station in the Niger River, where the observation records are questionable in a station in the Niger River."

In 5573-14: "Poorer hydrological performances (KGE'<0.5) are clustered in smaller tributaries in the arid area of South Africa and in a station in the Niger River, which is located downstream the Niger Inner delta. Therefore, the observed discharge has different characteristics which is not captured by the simulations."

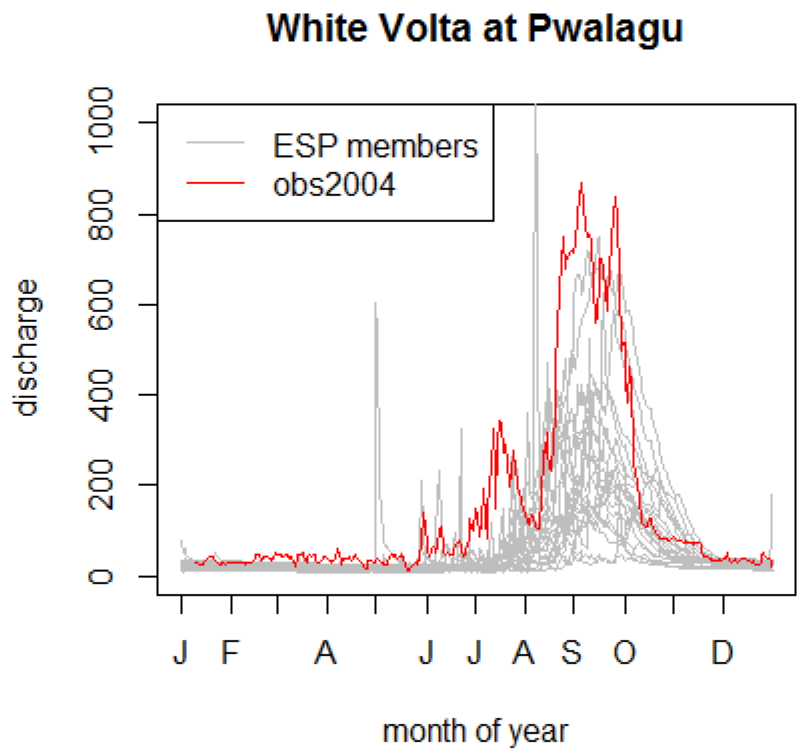
5574-25) The opposite of what?

The choice of word might not be clear; therefore we changed the sentence:

Decomposing the CRPSS for different regions in Africa shows that only a small number of stations in Eastern Africa (20 %) have skilful streamflow predictions, while in Western Africa the majority of stations (70-90 %) show skilful streamflow predictions.

5575-5) Why can it not be filtered out? You could simply run a conventional EPS as well. See main comment 1.

True. A conventional ESP has been derived from the long-term simulation (running LISFLOOD with the GPCP-corrected ERA-Interim over the time period 1989-2010). Calculating the Limit of predictability and the CRPSS over the 36 stations showed that those are slightly better than the skill received from AFFS. However, note, that it is not surprising as the conventional ESP is based on 20 years and not on only one year as the AFFS and hence, it reproduces the climatology better. Additional, five of the 20 years were used during calibration to optimize the performance of LISFLOOD, calculating a skill over the same time period and locations contribute to a skewed perspective on the results. However, regardless of those two objections, the main reason for not including the conventional EPS is that it is only to a limited amount relevant how well AFFS reproduces the general climatology, whereas AFFS capabilities to detect flood events is essential. Unfortunately, the locations of flood events don't coincide with the locations we have ground observations for; for which reason we scraped all kind of relevant flood-related information from several disaster data bases and based our evaluation of AFFS' flood forecasting capabilities on those. However, we have one location in the Volta in which there was a flood event reported and we also have ground observations for. For this location (see Figure below) the conventional ESP would have not been able to capture/forecast this flood event as it is outside of the range of the climatology build based on the 20-year long-term run. This was also confirmed by the limit of predictability and the CRPSS which are – for this particular location – much better for AFFS than for the ESP.



In any case, the skill (CRPSS) of the conventional ESP decreases similar to the skill of the AFFS with increasing lead time. Hence, the decrease in forecasting performance cannot only be associated solely due to possible inaccuracies of the ENS, but there must be other additional influencing factors – therefore we modified the part of the manuscript to:

~~"Whether the decrease in forecasting performance is caused by possibly inaccurate ENS cannot be assessed here, as the influence of the ENS cannot be filtered out.~~ As the skill of the conventional ESP (not shown here) decreases similar to the skill of the AFFS with increasing lead time, the decrease in forecasting performance cannot be affiliated to possible inaccuracies of the ENS only, but there must be other additional influencing factors. However, establishing the sources of predictability is beyond the scope of this paper, but subject of future research. However, […] "

5575-11) The relationship between historic model performance, theoretical skill and actual skill was analysed by Van Dijk et al. (doi:10.1002/wrcr.20251). Their conceptual framework can help improve your discussion.

Thank you for pointing us to this approach. We will consider this advice once we establish the source of predictability.

5576-8) Sounds very ad hoc, does it not cut both ways? Equally, there may be cases where you did not predict a flood while one occurred but was not reported.

True. But how can we speculate on something that we cannot even see any evidence of without being vague (not reported + not forecasted, BUT happened)? However, in order to not neglect this possibility, we are going to address this issue in the discussion:

The possibility that the database of observed events (Fig. 2) might not be complete allows also for the case that a flood event was neither forecasted nor reported but happened; which would result into a lower POD. However, there is no possibility to ascertain this issue unless more information become available.

5576-14) For these metrics to be interpretable they need to be compared to the performance of alternative methods, or is there an agreed 'standard'?

The contingency table together with the POD, FAR and CSI are standard measures for evaluating the performance of (individual) flood forecasting systems/approaches (meaning no comparison to other systems or approaches is required). As such they are widely used. Here only a minor number of citations (hundreds more can be found with a simple search in Google):

http://www.springer.com/environment/environmental+management/book/978-3-540-77852-3

http://www.srh.noaa.gov/srh/cwwd/hsd/verification/hydro/

http://www.opw.ie/hydrology/data/speeches/07%20-%20Lawless%20-%20New%20Standards%20and%20Methods%20for%20Validating%20Coastal%20Flood%20Forecasting%20Systems.pdf

http://www.chmi.cz/files/portal/docs/poboc/CB/pruvodce/vyhodnoceni_en.html

http://evidence.environment-agency.gov.uk/FCERM/Libraries/FCERM_Project_Documents/SCHO0305BIWA-E-E_pdf.sflb.ashx

http://webgr.irstea.fr/wp-content/uploads/2012/07/2010-WEEINK-MASTER.pdf

5576-17) How do I know that you are not cherry picking one of the more favourable examples? Include a few more (randomly selected) examples or leave this example out altogether.

We agree, it should have been mentioned that it is an example of a good performance of AFFS. We added this information to the manuscript to make it clear:

Figure 9 presents the flood forecast for the March 2003 event in the Save Basin (for location see Fig 1) as a visual example of a flood forecast obtained with AFFS. This flood forecast is one of the better ones achieved with AFFS. Note that there were no ground observations available

5577-21) I accept that, but floods of this size and duration are also more likely to lead to a phone call to authorities downstream, I would imagine, and might also be more likely picked up by remote sensing. You might counter argue that these early warning systems do not currently work well, but it is technically feasible to fix that. Pls discuss.

To "pick up" something with remote sensing means that it is happening in this moment or it already happened, whereas we are addressing here "forecasting", something that is not possible to be picked up by remote sensing as it hasn't happened yet; hence, those are 1-step-ahead information and very valuable to authorities dealing with flood mitigation.
The fact that AFFS can predict large-scale and long duration flood events better than small scale and short-duration ones is mostly related to the fact that these events are the result of large meteorological events, which are more likely to be predicted by the meteorological forecasts.
Of course if there is a flood event happening in the upper catchment it could be easily communicated with a simple phone call to the downstream areas. However, based on many information of different African authorities dealing with flood management issues we learned that this unfortunately does not always happen, particularly if the downstream catchment is covered by a different country. This has resulted into devastating consequences for the downstream areas in the past. One approach is of course improving the communication between countries, which might be challenging for different reasons. Another way of dealing with that is to establish a transnational/ continental flood forecasting system such as AFFS.

5578-2) Pls elaborate on the 'well in advance' aspect.

The exact number of days of lead-time varies for each flood event, and also for the same flood event depending on which critical threshold is being considered and the location within the catchment. For this reason it is not possible to assign one number for one event.
However, we modified the "well in advance" in the manuscript.

… to predict large-scale and long duration flood events several days in advance.

5578-7) As mentioned, presumably that cuts both ways.

See reply to "5576-8)" above.

5578-8 to 11) I appreciate that, but in that case why did you not consider that in your verification approach?

Yes, we agree. It would be very desirable to include this aspect into our analysis. To do so, we would need – among others – long-term ground measurements for the locations that floodings were reported. Unfortunately we usually don't have those for specific locations. Hence, with this limitation it is not possible for us to perform this analysis. However, once more observations become available we shall include this aspect into the analysis.

5578-15) First time this is raised, pls explain.

Valid point. The reason for not including it earlier i.e. in Section 5.2.2 / Table 4 is because it is not possible to make a static mask showing the areas that cover the boundaries of the ITCZ as those are not rigid. However, during the analysis we noticed the effect and got it confirmed by the study of Di. Giuseppe et al. (2013). Therefore we considered it worth mentioning in the discussion part.

The limitations of AFFS center around the detection of flood events with short durations (< week) and/or small affected areas (< 10 000km2)~~, as well as for flood events 15 occurring close to the boundaries of the Intertropical Convergence Zone~~. The difficulties in detecting relatively small and/or short duration flood events is most likely due to the combination of a) the limited precision given by SRFE (Satellite-based Rainfall Estimates) to capture small-scale meteorological events accurately in the correct time and place, and b) the relatively coarse grid size of 0.1×0.1° that AFFS is operating on, which might be too coarse for these type of floods. Also during the analysis it was noticed that flood events occurring close to the boundaries of the Intertropical Convergence Zone were not captured well (not shown within the analysis). ~~For flood events occurring closely at the boundaries of the ITCZ The~~ Forecasts in those areas may suffer from a displacement of the ITCZ controlling the onset and spatial extent of the West-Africa Monsoon, a conclusion also reached by Di Giuseppe et al.(2013).

5578-18) It would seem to me that a satellite rainfall based flood forecasting system (e.g. http://pmm.nasa.gov/node/187) can deal with this, although of course with very short (if any) lead time?

We agree with the self-raised objection given by the reviewer at the end of her/his statement. The system referred to (doi: 10.1029/2006WR005739) covers a different part of the whole chain of flood information systems ([seasonal forecasting →] medium-range forecasting → short-range forecasting → nowcasting → detection). Being based on "observations" (satellite-based rainfall estimations; TRMM 3B42) the system can only cover the detection to nowcasting of flood events, while the medium-range remains out of reach. Knowing the spatial resolution of TRMM 3B42 it is also questionable if the system is capable of detecting small-scale floodings. Any research in this direction is of course more than welcome.

5579-17) I don't think you have demonstrated this. You would have to first provide evidence that this system could/would lead to early warning beyond existing warning networks and improved upstream-downstream communication.

We don't agree with the reviewer on this point. Without doubt improving the upstream-downstream communication would be very valuable and could contribute to an enhanced preparedness in flood event situations. However, for many African countries this level of communication is unfortunately still – for various reasons – out of reach, and it is unknown "if" or "when" such agreements will be reached in future. However, flood events are real and occurring now and will continue to occur in dozens every year in Africa exposing challenges *to national and international aid organizations to cope with the resulting situations. Receiving medium-ranged flood forecast information opens a new door for improved preparedness independent of the level of communication between the countries of transboundary rivers.*

In this context, AFFS is an innovative system. Besides GloFAS, it is the first of its kind that is able to provide the whole African continent with probabilistic medium-range flood forecast information.

Slight modification to the text:

Concluding upon AFFS, this study has demonstrated that this system has a great potential to contribute to the reduction of flood-related losses in Africa by providing national and international aid organizations timely with medium-range ~~crucial~~ flood forecast information.